A Homologically Persistent Skeleton is a fast and robust descriptor of interest points in 2D images

Vitaliy Kurlin

vitaliy.kurlin@gmail.com, http://kurlin.org,
Microsoft Research Cambridge, 21 Station Road, Cambridge CB1 2FB, UK and
Department of Mathematical Sciences, Durham University, Durham DH1 3LE, UK

Abstract. 2D images often contain irregular salient features and interest points with non-integer coordinates. Our skeletonization problem for such a noisy sparse cloud is to summarize the topology of a given 2D cloud across all scales in the form of a graph, which can be used for combining local features into a more powerful object-wide descriptor.

We extend a classical Minimum Spanning Tree of a cloud to a Homologically Persistent Skeleton, which is scale-and-rotation invariant and depends only on the cloud without extra parameters. This graph

(1) is computable in time $O(n \log n)$ for any n points in the plane;

(2) has the minimum total length among all graphs that span a 2D cloud at any scale and also have most persistent 1-dimensional cycles;

(3) is geometrically stable for noisy samples around planar graphs.

Keywords: skeleton, Delaunay triangulation, persistent homology

1 Introduction: problem and overview

Pixel-based 2D images often contain *salient features* represented as points with non-integer coordinates. The resulting unstructured set is an example of a point *cloud* C, formally a finite metric space with pairwise distances between points.

The important problem in low level vision is to extract a meaningful structure from a given irregular cloud C. The traditional approach is to select a scale parameter, say a radius or the number of neighbors, and build a neighborhood graph. However, a real image may not have a single suitable scale parameter and we need to combine features found at multiple scales. This paper solves the skeletonization problem in its hardest form without any input parameters.

Parameterless skeletonization for sparse clouds. Given only an unstructured cloud $C \subset \mathbb{R}^2$ of points with any real coordinates, find a quickly computable structure that provably represents the topology of C at all scales.

Our solution is a 'homological' extension of a classical Minimum Spanning Tree $\mathrm{MST}(C)$ of a cloud C to a *Homologically Persistent Skeleton* $\mathrm{HoPeS}(C)$ that describes 1-dimensional cycles hidden in C over all possible scales α.

In section 2 we explain motivations for building HoPeS(C) and give a high level description of our contributions. In section 3 we compare our method with related work. In sections 4–5 we prove that HoPeS(C) or its subgraphs are

- **computable** in time $O(n \log n)$ for a cloud $C \subset \mathbb{R}^2$ of n points (Lemma 3)
- **invariant** up to rotations and uniform scale transformations (Lemma 4)
- **optimal** among all graphs capturing cycles of C at any scale (Theorem 5)
- **stable** under perturbations of samples C of graphs $G \subset \mathbb{R}^2$ (Corollary 8).

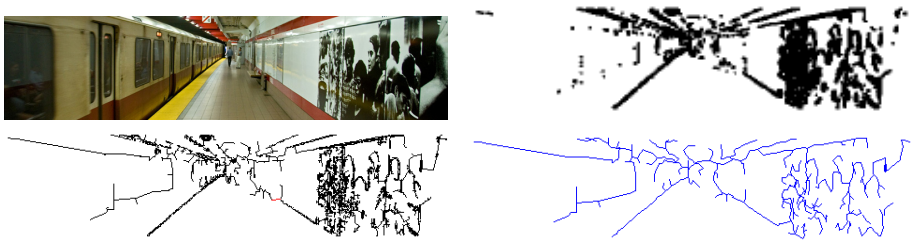

Fig. 1. Top: a cloud C of feature points. Bottom: HoPeS$'(C)$ and its simplification.

Fig. 1 shows the cloud C of $n = 7830$ feature points obtained by thresholding a real image in the top row, see details in section 6. The cloud C is the *only input* for producing the derived skeleton HoPeS$'(C)$ in the bottom row, where we kept only the most persistent cycle. The last picture of Fig. 1 is a simplified version of HoPeS$'(C)$ after removing short branches, see Definition 6. So HoPeS$'(C)$ provides a best 'guess' about the global topology of C in time $O(n \log n)$.

2 Our contributions and motivations of HoPeS(C)

Our parameterless skeletonization is based on *persistent homology*, which is the flagship method of Topological Data Analysis [10]. The key idea is to summarize topological features of data over all possible scales. A topological invariant that persists over a long interval of the scale is a true feature of the data, while noisy features have a short life span (a low persistence). The resulting persistent invariants are provably stable under noise, see [14, Appendix A]

Fig. 2 shows a cloud C on the integer lattice for simplicity, though our constructions work for any real coordinates. For any set $C \subset \mathbb{R}^2$ and $\alpha > 0$, the α-*offset* C^α consists of all points in $\mathbb{R}^2$ that are at most α away from C. Here α is the scale parameter (radius or width) of the α-offset $C^\alpha \subset \mathbb{R}^2$ around C.

We may gradually shrink a disk within itself to its center by making the radius smaller. We can not deform a circle to its center, because a smaller circle would be outside the original circle. So a circle is topologically non-trivial, while any closed loop in a disk is contractible. Spaces connected by such continuous deformations have the same *homotopy type*. We now formalize our problem.

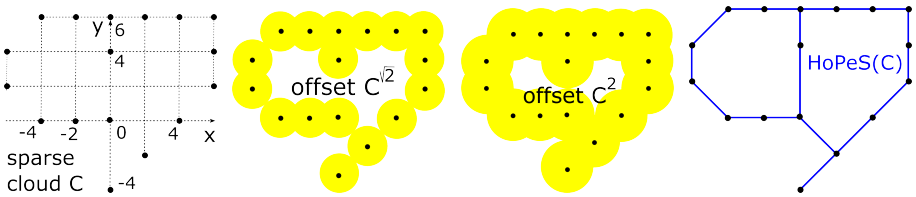

Fig. 2. A cloud C, α-offsets C^α and Homologically Persistent Skeleton HoPeS(C)

Multi-scale topological skeletonization: given a cloud $C \subset \mathbb{R}^2$, find a graph whose vertices are all points of C and whose suitable subgraphs have the homotopy type of the α-offset C^α for any α. A Homologically Persistent Skeleton HoPeS(C) is an optimal and stable skeleton satisfying the above requirements.

A cloud C is an *ε-sample* of (*ε-close* to) a graph $G \subset \mathbb{R}^2$ if $G \subset C^\varepsilon$ and $C \subset G^\varepsilon$. So any point of C is at most ε away from a point of G and any point of G is at most ε away from a point of C. The maximum possible value of ε is the upper bound of noise (the *Hausdorff* distance between G and its sample C).

Here is a high-level description of our contributions to skeletonization.

• Definition 2 introduces a Homologically Persistent Skeleton HoPeS(C) of a cloud $C \subset \mathbb{R}^2$ summarizing the persistence of 1-dimensional cycles in all C^α.

• Lemma 3 proves that, for a cloud $C \subset \mathbb{R}^2$ of any n points, HoPeS(C) has the size $O(n)$ and is computed in time $O(n \log n)$ without any extra parameters.

• Lemma 4 shows that HoPeS(C) is a scale-and-rotation invariant of $C \subset \mathbb{R}^2$.

• Theorem 5 proves that the reduced graph HoPeS($C;\alpha$) at any scale $\alpha > 0$ has the minimum length among all graphs that have the homotopy type of C^α.

• Theorem 7 guarantees that for any ε-sample of a simple enough graph $G \subset \mathbb{R}^2$, HoPeS$'$(C) is a correct topological reconstruction of G in the 2ε-offset $G^{2\varepsilon}$.

• Corollary 8 implies that the derived subgraph HoPeS$'$(C) is stable for any δ-perturbation of a cloud C that was ε-sampled around a planar graph G.

The novelty of this paper is not the fast algorithm for 1-dimensional persistence, but the new fundamental concept of a Homologically Persistent Skeleton HoPeS(C) that depends only a cloud $C \subset \mathbb{R}^2$ and solves the skeletonization problem without extra parameters and with guarantees in Theorems 5 and 7.

A graph without cycles is a *forest*. A connected forest is a *tree*. For a cloud $C \subset \mathbb{R}^2$, a *Minimum Spanning Tree* MST(C) is a tree that has the vertex set C and the minimum total length of edges, see Fig. 3. The *reduced forest* MST($C;\alpha$) is obtained from MST(C) by removing all open edges longer than 2α.

A connected graph G *spans* a cloud C if C is the vertex set of G. A graph G *spans* a possibly disconnected α-offset C^α if G has vertices at all points of the cloud C and any vertices of G are in the same connected component of G if and only if these vertices are in the same connected component of the α-offset C^α.

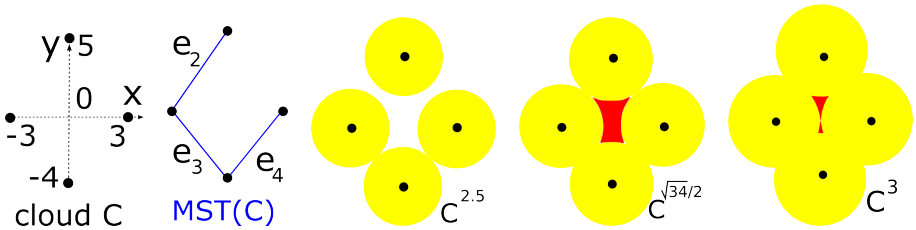

Fig. 3. Cloud C, minimum spanning tree MST(C) and α-offsets $C^{2.5}$, $C^{\sqrt{34}/2}$, C^3.

Points $p, q \in C$ are in the same *single-edge cluster* of C if $d(p,q) \leq 2\alpha$. Lemma 1 says that MST(C) is a universal optimal object that describes the 0-dimensional topology (all single-edge clusters) of C across all scales α.

Lemma 1 *For a cloud C and any scale $\alpha \geq 0$, the reduced forest* MST$(C;\alpha)$ *has the minimum total length of edges among all graphs that span C^α at the same scale α. Hence all connected components of the reduced forest* MST$(C;\alpha)$ *are in a 1-1 correspondence with all single-edge clusters of the cloud C.*

Lemma 1 and all later results are proved in [14, Appendix B]. Theorem 5 extends the optimality of MST(C) in Lemma 1 for clusters (dimension 0 approximation of C) to HoPeS(C) for cycles (dimension 1 approximation of C).

3 Comparison with related past skeletonization work

Our approach may look similar to the well-known scale-space theory [17] that suggests how to find a suitable scale. However, we do not choose any scale, we find topological features with longest life spans, which may not overlap. For instance, if one feature lives over the scale interval $1 \leq \alpha \leq 2$ and another over $3 \leq \alpha \leq 4$, then both features can not be captured at any fixed scale α. We can capture both features only by analyzing their life spans among all features.

The classical scale selection relies on analyzing data at discrete scales, usually proportional to powers of 2. The persistent homology works over the continuous scale so that all critical scales are found only from a given cloud, not by manually selecting a step size for incrementing the scale. Though we wouldn't say that persistent homology is 'perpendicular' to scale-space theory, our method is at least 'diagonal' to a scale selection, see diagonal gaps in Definiton 6.

To the best of our knowledge, all known skeletonization algorithms for clouds need extra parameters such as a scale α or a noise bound ε, e.g. [4]. Hence all these algorithms can not run on our minimal input, which is only a cloud C. Since a manual choice of parameters can be unfair, the experimental comparison with the past work seems impossible and we can compare only theoretical aspects.

N. Cornea et al. [6] stated the following requirements for skeletonization.

• *Topology*: a skeleton found from a noisy sample C is homeomorphic to (or has the homotopy type of) the original shape as in Theorem 7 from section 5.

• *Centering*: if a shape is well-sampled, a skeleton geometrically approximates the original shape in a small offset, see the 2ε-offset guarantee in Theorem 7.

• *Efficiency*: a near linear time in the number n of points as in our Lemma 3.

Our skeleton HoPeS(C) satisfies the extra conditions: independence of extra parameters, rotation-and-scale invariance and stability under bounded noise.

R. Singh et al. [18] approximated a skeleton of a shape by a subgraph of a Delaunay triangulation based on 2nd order Voronoi regions. The algorithm has 3 threshold parameters: K for the minimum number of edges in a cycle and $\delta_{min}, \delta_{max}$ for inserting and merging Voronoi regions. M. Aanjaneya et al. [1] solved a related problem approximating a metric on a large input graph by a metric on a small output graph. So the input is a graph, not a cloud of points.

Starting from a noisy sample of an unknown graph G with a scale parameter, X. Ge et al. [11] produced the Reeb graph with the same number of loops as the graph G. This output is an abstract graph of a simplicial complex on a cloud C and is not intrinsically embedded into any space even if $C \subset \mathbb{R}^2$. [11, section 3.3] reported 'spurious branches or loops in the Reeb graph constructed no matter how we choose a radius or a number of neighbours to decide the scale'.

F. Chazal et al. [3] introduced a new α-Reeb graph for a graph reconstruction in different settings. The distance between points in a noisy sample C is measured geodesically within a given neighborhood graph on C, while we consider offsets of C with respect to the ambient distance in $\mathbb{R}^2$. Their algorithm has the same fast time $O(n \log n)$ and they also gave conditions when the reconstructed graph has a required homotopy type.

T. Dey et al. [8] built a complex depending on a user-defined graph that spans a cloud C of n points. This Graph Induced Complex GIC has the same homology H_1 as the Rips complex of a cloud C at a suitable scale α. The 2D skeleton of GIC needed for computing H_1 has the size $O(n^3)$ in a worst case.

For image segmentation, α-offsets were similarly used in [16] (with 2 extra parameters) and in [13] (without parameters). A Homologically Persistent Skeleton can be defined for arbitrary filtrations on a cloud in any metric space [15].

Papers	[11], 2011	[1], 2012	[8], 2013	[3], 2014	this paper
Extra input	radius r	radius r, noise ε	graph spanning C	scale α	no parameters
Complexity	$O(n \log n)$	$O(n^2)$	at least $O(n^3)$	$O(n \log n)$	$O(n \log n)$

Table 1. Comparison of similar skeletonization methods for unstructured clouds.

The discussion in [2, section 13] proposed to select features by persistence [10] and led us to the new concept of HoPeS(C) in Definition 2. The key advantage of our approach over the past work is the *absence of any user-defined parameters.*

• HoPeS(C) of a cloud C has no extra input parameters (such as ε or α) that are needed in all past skeletonization algorithms for an unstructured cloud C.

• For a cloud $C \subset \mathbb{R}^2$ of any n points, the skeleton HoPeS(C) with $O(n)$ edges can be found in time $O(n \log n)$, which is comparable only with [3], [11], [18].

• HoPeS(C) is the *first universal structure* on a cloud C that summarizes all cycles of C^α and has a subgraph HoPeS$'(C)$ stable under perturbations of C.

• HoPeS$'(C)$ for an ε-sample C of G approximates G in the thin offset $G^{2\varepsilon} \subset \mathbb{R}^2$. [3] has guarantees for the abstract Gromov-Hausdorff distance, not in $\mathbb{R}^2$.

• Theorem 5 gives guarantees only in simple terms of a graph $G \subset \mathbb{R}^2$ and its noisy ε-sample $C \subset \mathbb{R}^2$, while [11, Theorem 3.1] needs a complex K with a homotopy equivalence $h : K \to G$ that ε-approximates the metrics of K and G.

4 A Homologically Persistent Skeleton and its optimality

Here we give a rather intuitive introduction into homology theory using only α-offsets C^α as typical spaces, see rigorous definitions in [14, Appendix A].

The *0-dimensional homology* H_0 counts connected components. Formally, $H_0(C^\alpha)$ is the group (or vector space of linear combinations with coeffiecients in $\mathbb{Z}_2 = \{0, 1\}$) generated by the components of C^α. For instance, the offset $C^{2.5}$ in Fig. 3 has 2 components. Hence $H_0(C^{2.5}) = \mathbb{Z}_2 \oplus \mathbb{Z}_2$ has rank (*dimension*) 2.

The *1-dimensional homology* H_1 of $C^\alpha \subset \mathbb{R}^2$ similarly counts *holes* in C^α (bounded regions in the complement $\mathbb{R}^2 - C^\alpha$). For example, the offset $C^{\sqrt{34}/2}$ in Fig. 3 has 1 red hole, so $H_1(C^{\sqrt{34}/2}) = \mathbb{Z}_2$. This hole splits into 2 holes at $\alpha = 3$, hence $H_1(C^3) = \mathbb{Z}_2 \oplus \mathbb{Z}_2$. The smaller of the 2 holes disappears when $\alpha = \frac{25}{8}$ is the circumradius of the triangle on vertices $(\pm 3, 0)$ and $(0, -4)$, so $H_1(C^{25/8}) = \mathbb{Z}_2$. The remaining hole dies when $\alpha = \frac{17}{5}$ is the circumradius of the triangle on vertices $(\pm 3, 0)$ and $(0, 5)$, hence $H_1(C^{17/5}) = 0$ is trivial.

All α-offsets form an ascending *filtration* (a nested sequence of spaces) $C = C^0 \subset \ldots \subset C^\alpha \subset \ldots \subset C^{+\infty} = \mathbb{R}^2$. These inclusions induce linear maps in H_1:

$$C^{2.5} \subset C^{\sqrt{34}/2} \subset C^3 \subset C^{25/8} \subset C^{17/5} \text{ induce } 0 \to \mathbb{Z}_2 \to \mathbb{Z}_2 \oplus \mathbb{Z}_2 \to \mathbb{Z}_2 \to 0.$$

The sequence of the linear maps in H_1 above splits into 2 simpler sequences:

hole 1 lives over the interval $\frac{\sqrt{34}}{2} \le \alpha < \frac{17}{5}$, namely $0 \to \mathbb{Z}_2 \to \mathbb{Z}_2 \to \mathbb{Z}_2 \to 0$,

hole 2 lives over the short interval $3 \le \alpha < \frac{25}{8}$, namely $0 \to 0 \to \mathbb{Z}_2 \to 0 \to 0$.

At $\alpha = 3$ when the initial hole splits into 2 smaller holes, we assume that one of the holes 'inherits' (continues the life of) the previous hole, while another hole is 'newborn' at the splitting moment. The standard convention is to give preference to a longer living hole. So the life spans (the *barcode*) of the filtration $\{C^\alpha\}$ are $[\frac{\sqrt{34}}{2}, \frac{17}{5})$ and $[3, \frac{25}{8})$. We plot the endpoints of these bars as red dots with coordinates $(\text{birth}, \text{death})$ in the *persistence diagram* PD$\{C^\alpha\}$, see Fig. 4.

This diagram is a summary of life spans of holes (1-dimensional homology classes) of C^α across all scales α. The key result of persistent homology is the Stability Theorem [5] roughly saying that any small perturbation of the cloud C gives rise to a similar small perturbation of the diagram $\mathrm{PD}\{C^\alpha\}$ in the plane.

If a hole of C^α is born, then this hole becomes enclosed by a cycle through points of C. The last longest edge in this enclosing cycle is added at the *birth* time α of the hole and is *critical* for the hole in question. Hole 1 born at $\alpha = \frac{\sqrt{34}}{2}$ has the critical edge e_1, see Fig. 4. Hole 2 born at $\alpha = 3$ has the critical edge e_5.

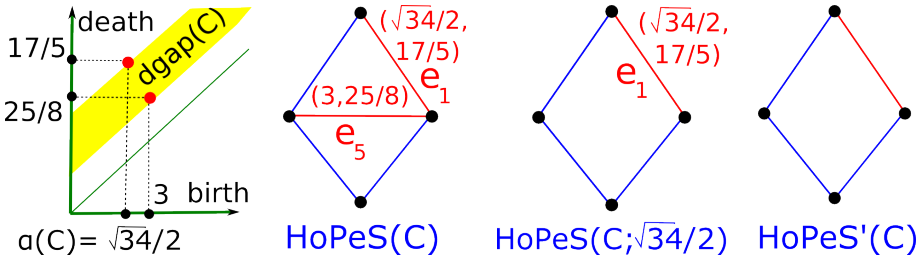

Fig. 4. Diagram $\mathrm{PD}\{C^\alpha\}$ for the cloud C in Fig. 3 and skeletons from Definitions 2, 6.

For any filtration $\{C^\alpha\}$, each red dot in $\mathrm{PD}\{C^\alpha\}$ has a corresponding critical edge e (between points of C) with the label $(\mathrm{birth}(e), \mathrm{death}(e))$. Our Definition 2 transforms the diagram $\mathrm{PD}\{C^\alpha\}$ of disconnected points into a universal structure on the data cloud C summarizing the persistence of holes in $\{C^\alpha\}$ for all α.

Definition 2 *For a cloud C, a* Homologically Persistent Skeleton $\mathrm{HoPeS}(C)$ *is the union of* $\mathrm{MST}(C)$ *and all critical edges with their labels* $(\mathrm{birth}, \mathrm{death})$, *see Fig. 4. The* reduced *skeleton* $\mathrm{HoPeS}(C;\alpha)$ *is obtained from* $\mathrm{HoPeS}(C)$ *by removing all edges longer than* 2α *and all critical edges e with* $\mathrm{death}(e) \le \alpha$.

If $\alpha = 0$, then $\mathrm{HoPeS}(C;0) = C$ is the given cloud. By Definition 2 a critical edge e belongs to the reduced skeleton $\mathrm{HoPeS}(C;\alpha)$ if and only if $\mathrm{birth}(e) \le \alpha < \mathrm{death}(e)$. So a critical edge e is added to $\mathrm{HoPeS}(C;\alpha)$ at $\alpha = \mathrm{birth}(e)$ and is later removed at the larger scale $\alpha = \mathrm{death}(e)$. The cloud C in Fig. 3 has $\mathrm{HoPeS}(C;\frac{\sqrt{34}}{2}) = \mathrm{MST}(C) \cup e_1$, but $\mathrm{HoPeS}(C;3)$ coincides with $\mathrm{HoPeS}(C)$.

The filtration $\{\mathrm{HoPeS}(C;\alpha)\}$ may not be monotone with respect to the scale α. But if $\mathrm{HoPeS}(C;\alpha)$ has become connected, it will stay connected for all larger α. Indeed, removing a critical edge destroys only a cycle, not connectivity.

Similarly to $\mathrm{MST}(C)$, a Homologically Persistent Skeleton $\mathrm{HoPeS}(C)$ is unique in a general position when the distances between all points of C are different.

Lemma 3 *For any cloud $C \subset \mathbb{R}^2$ of n points, a Homologically Persistent Skeleton* $\mathrm{HoPeS}(C)$ *has the size $O(n)$ and is computable in time $O(n \log n)$.*

Lemma 4 below help visualize the 1-dimensional persistence diagram $\mathrm{PD}\{C^\alpha\}$ directly on the cloud C. Lemma 4 justifies that $\mathrm{HoPeS}(C)$ is suitable for Computer Vision applications where a scale-and-rotation invariance is important.

Lemma 4 *For a cloud $C \subset \mathbb{R}^2$, the 1-dimensional persistence diagram $\mathrm{PD}\{C^\alpha\}$ of the filtration of α-offsets C^α can be reconstructed from a Homologically Persistent Skeleton $\mathrm{HoPeS}(C)$. The topological structure of $\mathrm{HoPeS}(C)$ is invariant under any affine transformation whose 2×2 matrix has equal eigenvalues.*

Our first main Theorem 5 says that $\mathrm{HoPeS}(C)$ is an optimal graph that extends $\mathrm{MST}(C)$ and captures the persistence of all holes in the filtration $\{C^\alpha\}$.

Theorem 5 *For any cloud $C \subset \mathbb{R}^2$ and any $\alpha > 0$, the graph $\mathrm{HoPeS}(C; \alpha)$ has the minimum total length of edges over all graphs $G \subset C^\alpha$ that span the α-offset C^α and induce an isomorphism in 1-dimensional homology $H_1(G) \to H_1(C^\alpha)$.*

A graph G *spans* C^α if $G \subset C^\alpha$ induces an isomorphism $H_0(G) \cong H_0(C^\alpha)$. An isomorphism $H_1(G) \cong H_1(C^\alpha)$ means that the graph G has the homotopy type of the α-offset $C^\alpha \subset \mathbb{R}^2$. Hence our Homologically Persistent Skeleton $G = \mathrm{HoPeS}(C)$ solves the multi-scale skeletonization problem stated in sections 1–2.

5 The reconstruction theorem and stability of HoPeS(C)

A Homologically Persistent Skeleton $\mathrm{HoPeS}(C)$ contains all 1-dimensional cycles in the offsets C^α across the full range of α. It is natural to select cycles with highest persistence to get a smaller subgraph $\mathrm{HoPeS}'(C) \subset \mathrm{HoPeS}(C)$. So we select not a scale as in scale-space theory, but a widest diagonal gap in the persistence diagram $\mathrm{PD}\{C^\alpha\}$. This widest gap makes sense for finite sets C and for any *compact* set $S \subset \mathbb{R}^2$ that is a finite union of closed topological disks.

Definition 6 *For a compact set $S \subset \mathbb{R}^2$ and the ascending filtration of offsets S^α, a* diagonal gap *in the persistence diagram $\mathrm{PD}\{S^\alpha\}$ is a largest (by inclusion) strip $\{0 \le a < y - x < b\}$ that has no points from the diagram, see Fig. 3.*

The widest diagonal *gap* $\mathrm{dgap}(S)$ *has the largest width* $|\mathrm{dgap}(S)| = b - a$. *Let the subdiagram $\mathrm{PD}'\{S^\alpha\} \subset \mathrm{PD}\{S^\alpha\}$ have only the points above* $\mathrm{dgap}(S)$. *The* critical *scale $\alpha(S)$ is the maximum* birth *over all* $(\mathrm{birth}, \mathrm{death}) \in \mathrm{PD}'\{S^\alpha\}$.

For a cloud $C = S$, the derived *skeleton $\mathrm{HoPeS}'(C)$ is obtained from $\mathrm{HoPeS}(C)$ by removing (1) all edges longer than $2\alpha(C)$, and (2) all critical edges either with* $\mathrm{death} \le \alpha(C)$ *or with* $(\mathrm{birth}, \mathrm{death})$ *below the widest diagonal gap* $\mathrm{dgap}(C)$.

In Definition 6 if there are different gaps with the same width, we say that the gap with largest values along the vertical death axis has the largest width. The cloud C in Fig. 3 has the widest gap $\mathrm{dgap}(C)$ between the points $(\frac{\sqrt{34}}{2}, \frac{17}{5})$ and $(3, \frac{25}{8})$ in $\mathrm{PD}\{C(\alpha)\}$, so the critical scale is $\alpha(C) = \frac{\sqrt{34}}{2}$, see Fig. 4.

Condition (1) above guarantees that $\mathrm{HoPeS}'(C) \subset \mathrm{HoPeS}(C; \alpha(C))$, because all long critical edges e with $\mathrm{birth}(e) > \alpha(C)$ are removed, see [14, Appendix B]. Condition (2) filters out cycles with early deaths and low persistence, but $\mathrm{HoPeS}'(C) \neq \mathrm{HoPeS}(C; \alpha(C))$. Instead of selecting a fixed scale as in scale-space theory, we select cycles by their persistence across all scales α.

We define concepts needed for Theorem 7. A non-self-intersecting cycle L in a graph $G \subset \mathbb{R}^2$ is *basic* if L encloses a bounded region of $\mathbb{R}^2 - G$. When α is increasing, the hole enclosed by the α-offset L^α is born at $\alpha = 0$ and dies at the scale $\alpha = \rho(L)$ that is called the *radius* of the cycle L. So the initial hole enclosed by L has the life span $[0, \rho(L))$. The heart-shaped hole in the first picture of Fig. 5 completely dies at $\alpha = \rho(L)$, which holds for any convex hole.

In general, when α is increasing new holes can be born in G^α, let they be enclosed by $L_1, \dots, L_k$ at their birth times. The *thickness* $\theta(G) = \max\limits_{j=1,\dots,k} \rho(L_j)$ is the maximum persistence of these smaller holes born during the evolution of offsets G^α. If no such holes appear, then $\theta = 0$, otherwise $\theta > 0$, see Fig. 5.

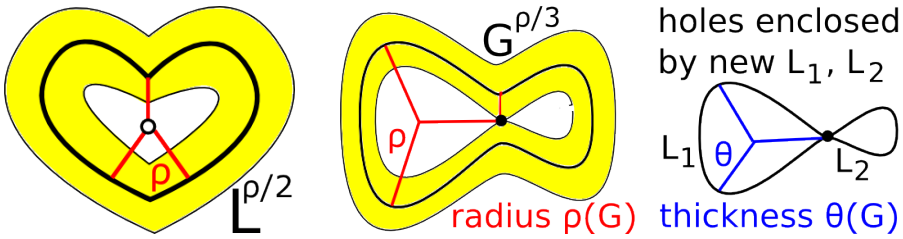

Fig. 5. The 'heart' graph has thickness $\theta = 0$. The 'figure-eight' graph has $\theta > 0$.

Theorem 7 says that $\mathrm{HoPeS}'(C)$ is a close approximation to a graph G from any its ε-sample C. The homotopy type of a connected graph G is determined by its $H_1(G)$. Namely, G continuously deforms to a wedge of $\dim H_1(G)$ loops.

Theorem 7 *Let C be any ε-sample of a connected graph $G \subset \mathbb{R}^2$ with a thickness $\theta(G) \geq 0$ and $m \geq 1$ basic cycles having ordered radii $\rho_1 \leq \ldots \leq \rho_m$. If $\rho_1 > 7\varepsilon + \theta(G) + \max\limits_{i=1,\dots,m-1} \{\rho_{i+1} - \rho_i\}$, then the critical scale $\alpha(C) \leq \varepsilon$, and the derived skeleton $\mathrm{HoPeS}'(C)$ is 2ε-close to G and has the homotopy type of G.*

The inequality above means that the cycles of the graph G have 'comparable' sizes, i.e. the smallest radius ρ_1 is larger by a good margin than any gap $\rho_{i+1} - \rho_i$ between the ordered radii. Hence the diagonal gap $\{\theta(G) < \mathrm{death} - \mathrm{birth} < \rho_1\}$ in the diagram $\mathrm{PD}\{G^\alpha\}$ of the graph G will remain wide enough to be automatically recognized in the perturbed diagram $\mathrm{PD}\{C^\alpha\}$ for any ε-sample C of G.

Theorem 7 is stronger than any estimate of homology from noisy samples. In addition we build on a sample C an actual skeleton $\mathrm{HoPeS}'(C)$ that is 2ε-close to an unknown graph G. Theorem 7 extends simpler [13, Theorem 32], which works only for a much smaller class of graphs $G \subset \mathbb{R}^2$ with thickness $\theta = 0$.

Corollary 8 *In the conditions of Theorem 7 if another cloud $\tilde{C}$ is δ-close to C, then the perturbed derived skeleton* $\mathrm{HoPeS}'(\tilde{C})$ *is* $(2\delta + 4\varepsilon)$*-close to* $\mathrm{HoPeS}'(C)$.

We can't expect that $\mathrm{HoPeS}'(C)$ is locally stable for any cloud C, because a minimum spanning tree $\mathrm{MST}(C)$ is sensitive to perturbations of C. However, Corollary 8 guarantees the overall stability of the derived skeleton (within a small offset) in the most practical case for noisy sample of graphs.

6 Algorithm, experiments and practical applications

[14, Appendix A] justifies that complicated α-offsets C^α can be replaced by simpler α-complexes $C(\alpha)$, which filter a Delaunay triangulation $\mathrm{Del}(C)$. Starting from a cloud $C \subset \mathbb{R}^2$ of n points, we build $\mathrm{Del}(C)$ in time $O(n \log n)$ with $O(n)$ space. Regions in the complement $\mathbb{R}^2 - C(\alpha)$ are dual to their boundaries. This duality [10] reduces 1-dimensional persistence of cycles in the filtration $\{C(\alpha)\}$ to 0-dimensional persistence of connected components in $\mathbb{R}^2 - C(\alpha)$.

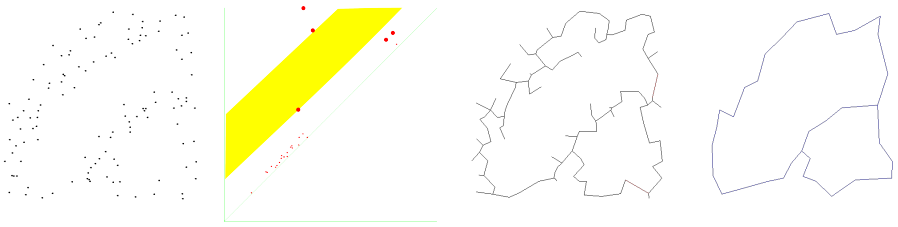

Fig. 6. A sample C of O45, diagram $\mathrm{PD}\{C(\alpha)\}$, $\mathrm{HoPeS}'(C)$ and its simplification.

The 0-dimensional persistence is computed in time $O(nA^{-1}(n))$ using a union-find structure [10], where $A^{-1}(n)$ is the slow growing inverse Ackermann function. We extend this algorithm by recording a critical edge along which regions of $\mathbb{R}^2 - C(\alpha)$ merge when α is decreasing, see [14, Appendix C].

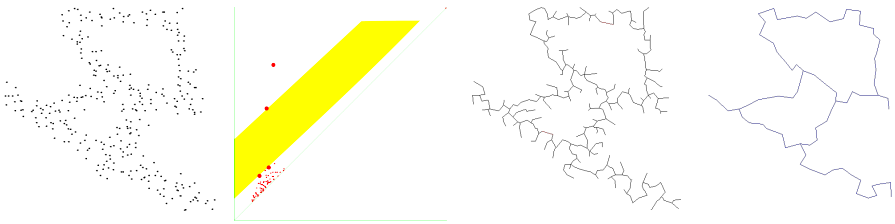

Fig. 7. A sample C of D33, diagram $\mathrm{PD}\{C(\alpha)\}$, $\mathrm{HoPeS}'(C)$ and its simplification.

Fig. 6 shows 121 random points sampled from a real image of hieroglyph O45. The second picture of Fig. 6 is the diagram PD$\{C(\alpha)\}$ with a widest diagonal gap clearly separating the noise near the diagonal from 2 red points corresponding to 2 cycles in the derived graph HoPeS$'(C)$. Theorem 7 gives the lower bound $\alpha(C)$ for the unknown noise level ε. We use this intrinsic critical scale $\alpha(C)$ for pruning short branches and collapsing short edges to get a simplified version of HoPeS$'(C)$ in the last picture of Fig. 6, see details in [14, Appendix C]. Fig. 7 has similar results for 321 points sampled from another hieroglyph D33.

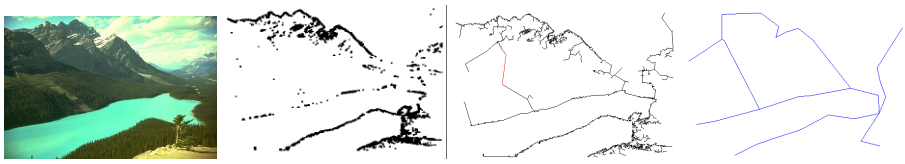

Fig. 8. Image BSD176035, cloud C of 3603 points, HoPeS$'(C)$ and its simplification.

In Fig. 8 we selected feature points from a challenging image by simply comparing the color of each pixel with the average in 5×5 neighborhood. The threshold for the normal deviation in the 3-dimensional RGB space was 65 as in Fig. 1. Fig. 9 shows similar results for the normal deviation 100 of the color.

Fig. 9. Image BSD42049, cloud C of 1763 points, HoPeS$'(C)$ and its simplification.

Table 2 has the running time in milliseconds for the database BSD500, where all images have 481×321 pixels and we used 2 thresholds in each image, see details in [14, Appendix D]. We used a small laptop with 1.33GHz RAM 2GB to show that the algorithm is fast for embedded systems. The C++ code is on author's page `http://kurlin.org/projects/persistent-skeletons.php`.

Images from BSD500	42049	42049	176035	176035	175083	175083	134049	134049
Time for a cloud C, ms	1022	1336	1017	1038	1066	1015	1051	1165
Points in the cloud C	2664	3604	3603	4249	3928	4950	4396	6767
Time for HoPeS$'(C)$, ms	969	1789	1898	2629	2143	3602	3780	6259

Table 2. Time for extracting C from images in BSD500 and computing HoPeS$'(C)$.

We have demonstrated the following practical applications of HoPeS(C).

• Robust recognition of low quality scans in Fig. 6, 7, see more results in [14]. Such visual markers [7] can replace shop barcodes not readable by humans.

• A fast topological summary of images, see Fig. 1, 8 and more details in [14].

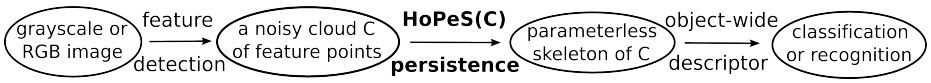

Fig. 10. Pipeline for building an object-wide descriptor from noisy local features

References

1. M. Aanjaneya, F. Chazal, D. Chen, M. Glisse, L. Guibas, D. Morozov. Metric graph reconstruction from noisy data. *IJCGA*, v. 22 (2012), 305–325.
2. D. Attali, J.-D. Boissonnat, H. Edelsbrunner. Stability and computation of medial axes – a state-of-the-art report. In *Math. Foundations of Visualization, Computer Graphics, and Massive Data Exploration*, 109-125. Springer, 2009.
3. F. Chazal, R. Huang, J. Sun. Gromov-Hausdorff approximation of filament structure using Reeb-type graph. *Discrete Comp. Geometry*, v. 53 (2015), 621–649.
4. A. Chernov, V. Kurlin. Reconstructing persistent graph structures from noisy images, *Image-A*, v. 3 (2013), p. 19-22.
5. D. Cohen-Steiner, H. Edelsbrunner, J. Harer. Stability of persistence diagrams. *Discrete and Computational Geometry*, v. 37 (2007), p. 103–130.
6. N. Cornea, D. Silver, and P. Min. Curve-Skeleton Properties, Applications, and Algorithms *IEEE Trans. Visualization Comp. Graphics*, v. 13 (2007), 530–548.
7. E. Costanza and J. Huang. Designable visual markers. *Proceedings of SIGCHI 2009* : Special Interest Group on Computer-Human Interaction, p. 1879-1888.
8. T. Dey, F. Fan, and Y. Wang. Graph induced complex on data points. *Proceedings of SoCG 2013* : Symposium on Computational Geometry, p. 107–116.
9. H. Edelsbrunner. The union of balls and its dual shape. *Discrete Computational Geometry*, v. 13 (1995), p. 415–440.
10. H. Edelsbrunner, J. Harer. *Computational topology. An introduction.* AMS, 2010.
11. X. Ge, I. Safa, M. Belkin, and Y. Wang. Data skeletonization via Reeb graphs. *Proceedings of NIPS 2011*, p. 837–845.
12. V. Kurlin. A fast and robust algorithm to count topologically persistent holes in noisy clouds. *Proceedings of CVPR 2014*, p. 1458–1463.
13. V. Kurlin. Auto-completion of contours in sketches, maps and sparse 2D images based on topological persistence. *Proceedings of CTIC 2014*, p. 594–601.
14. V. Kurlin. A Homologically Persistent Skeleton is a fast and robust descriptor of interest points in 2D images (full version of this paper at `http://kurlin.org`).
15. V. Kurlin. A one-dimensional Homologically Persistent Skeleton of an unstructured point cloud in any metric space. *Computer Graphics Forum*, to appear (2015). Available at `http://kurlin.org/projects/persistent-skeleton.pdf`.
16. D. Letscher, J. Fritts. Image segmentation using topological persistence. *Proceedings of CAIP 2007* : Computer Analysis of Images and Patterns, 587–595.
17. T. Lindeberg. *Scale-Space Theory in Computer Vision*, Kluwer Publishers, 1994.
18. R. Singh, V. Cherkassky, N. Papanikolopoulos. Self-organizing maps for the skeletonization of sparse shapes. *Tran. Neural Networks*, v. 11 (2000), 241–248.